Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes

Gia-Fen T.Chen and Masayori Inouye
Department of Biochemistry, Robert Wood Johnson Medical School – UMDNJ at Rutgers, Piscataway, NJ 08854, USA

Received December 11, 1989; Revised and Accepted February 11, 1990

ABSTRACT

AGA and AGG codons for arginine are the least used codons in *Escherichia coli*, which are encoded by a rare tRNA, the product of the *dnaY* gene. We examined the positions of arginine residues encoded by AGA/AGG codons in 678 *E. coli* proteins. It was found that AGA/AGG codons appear much more frequently within the first 25 codons. This tendency becomes more significant in those proteins containing only one AGA or AGG codon. Other minor codons such as CUA, UCA, AGU, ACA, GGA, CCC and AUA are also found to be preferentially used within the first 25 codons.

The effects of the AGG codon on gene expression were examined by inserting one to five AGG codons after the 10th codon from the initiation codon of the *lacZ* gene. The production of β-galactosidase decreased as more AGG codons were inserted. With five AGG codons, the production of β-galactosidase (Gal-AGG5) completely ceased after a mid-log phase of cell growth. After 22 hr induction of the *lacZ* gene, the overall production of Gal-AGG5 was 11% of the control production (no insertion of arginine codons). When five CGU codons, the major arginine codon were inserted instead of AGG, the production of β-galactosidase (Gal-CGU5) continued even after stationary phase and the overall production was 66% of the control. The negative effect of the AGG codons on the Gal-AGG5 production was found to be dependent upon the distance between the site of the AGG codons and the initiation codon. As the distance was increased by inserting extra sequences between the two codons, the production of Gal-AGG5 increased almost linearly up to 8 fold. From these results, we propose that the position of the minor codons in an mRNA plays an important role in the regulation of gene expression possibly by modulating the stability of the initiation complex for protein synthesis.

INTRODUCTION

Non-random usage of synonymous codons has been clearly demonstrated in both prokaryotes and eukaryotes (for example, see refs. 1–3). In *Escherichia coli*, codon usage for the major outer membrane lipoprotein, the most abundant protein in *E. coli*, has been shown to be extremely biased; less than half of the available codons are used for the lipoprotein (4,5). Similar biased codon usage was also shown in *E. coli* ribosomal proteins, and such non-random codon usage was speculated to be due to preferential usage of major isoaccepting species of tRNA (6,7). From the analysis of codon usage in various *E. coli* proteins and relative quantities of tRNAs, a strong positive correlation between tRNA content and the occurrence of respective codons was demonstrated, and it was proposed that *E. coli* genes encoding abundant protein species selectively use the 'optimal codon' or major codons as determined by the abundance of isoaccepting tRNAs (2).

Because of these observations it is generally believed that the protein production from a gene containing minor codons or non-optimal codons is less efficient than that from a gene containing no minor codons. In contrast to this notion, however, some researchers have shown that minor codons are not used to modulate the rate of protein production (8,9). Recently, it was demonstrated that there is a difference in translation rate between common codons and minor codons (10). However, the difference in overall translation time between the genes with and without minor codons was small, and therefore it appears that the minor codon usage does not contribute to significant effects on gene expression at the level of translation (10,11). On the other hand, in *Saccharomyces cerevisiae* it has been demonstrated that replacing an increasing number of major codons with synonymous minor codons at the 5′ end of the coding sequence of the gene for phosphoglycerate kinase caused a dramatic decrease of the gene expression (12). In this work, the authors concluded that efficient mRNA translation is required for maintaining mRNA stability in *S. cerevisiae*.

In this report, we first examine the position of AGA/AGG codons and other minor codons in *E. coli* genes, and found that in the genes containing AGA/AGG codons and other minor codons, they are used in a significantly higher frequency within the first 25 codons. Next we attempted to investigate the role of the minor codons in gene expression in *E coli*. For this purpose we chose the AGG codon for arginine for the following reasons: (a) AGG for arginine is the least used codon among the codons

used in *E coli*. Arginine can be coded for by six codons; CGU, CGC, CGA, CGG, AGA and AGG. In 199 *E. coli* genes examined for codon usage (3) arginine codons are used a total of 4,149 times. Out of these codons CGU, CGC, CGA, CGG, AGA and AGG are used 2,047 times (49.3%), 1,584 (38.2%), 161 (3.8%), 250 (6.0%), 69 (1.7%) and 38 (0.9%), respectively. (b) AGA and AGG codons are recognized by a very rare tRNA (2), and the tRNA gene for these two codons was identified to be *dnaY*, indicating that this rare tRNA possibly plays a role in essential cellular functions (13). (c) In the *E. coli* genes containing AGA/AGG codons, these codons are preferentially used within the first 25 codons. We found that even if five AGG codons were tandemly inserted after the 10th codon from the initiation codon of the *lacZ* gene, the rate of the production of β-galactosidase (Gal-AGG5) was not significantly different up to mid-log phase of cell growth from that of β-galactosidase (Gal-CGU5) in which all the AGG codons in Gal-AGG5 were replaced with CGU, the major arginine codon. However, after mid-log phase, the Gal-AGG5 production completely ceased, while the Gal-CUG5 production continued even in stationary phase. The negative effect of the AGG codons on the *lacZ* expression was, however, completely suppressed by increasing the spacing between the AGG insertion site and the initiation codon. Our data indicate that the number and the position of the arginine minor codons in an mRNA play an important role in modulating gene expression at the level of translation, in particular, in later growth phase. A possible mechanism will be proposed as to how the negative effect of minor codons in an mRNA or its expression can be suppressed when their position in the mRNA is kept far from the initiation codon. The present finding together with the fact that all minor codons are preferentially used in the first 25 codons in *E. coli* genes presents important implication of general roles of minor codons in gene expression.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

E. coli strain SB4288 K-12 F$^-$ *recA thi*-1 *relA mal*-24 *spc*12 *supE*-50 *DE5* Δ(*lac-proB*) (14) and JM83 *ara* Δ(*lac-proAB*) *rpsL* φ80 *lacZ* ΔM15 *relA*$^+$ (15) were used.

The plasmids pIT-lac1-CGU5 and pIT-lac1-AGG1, -AGG2, -AGG3, -AGG4 and -AGG5 were constructed from pIT-lac1 previously designated as pIN-III [*lpp* $^{p-5}$ *lacZ* (16)] as described in the text. The plasmids were screened by DNA sequencing (17). Plasmids containing different numbers of the linker shown in Figure 2C were constructed as follows: pIT-lac1-AGG5 was digested with *Hin*dIII, followed by treatment with bacterial alkaline phosphatase. Before ligation of the linker, the oligonucleotides were annealed and ligated by themselves to form various multimers. The mixture of these multimers was then used for ligation at the *Hin*dIII site of pIT-lac1-AGG5. The recombinant DNA procedures were carried out according to the methods described by Maniatis et al. (18).

Protein Analysis

Cells harboring various plasmids were grown at 37°C in M9-Casamino Acid medium (19). At a Klett unit of 15, IPTG was added to a final concentration of 2mM. Aliquots were taken at time intervals after the addition of IPTG. The production of β-galactosidase was analyzed by SDS-polyacrylamide gel electrophoresis as described previously (16). The gels were stained by Coomassie Brilliant Blue and the production of β-galactosidase was quantitated by scanning the stained gels with a GS300 densitometer from Hoeffer Scientific Instruments.

Enzymes and Chemicals

Restriction endonucleases were obtained from New England Biolabs. IPTG was purchased from Sigma Chemical Co. The DNA oligomers were synthesized on an Applied Biosystems 380B DNA Synthesizer.

RESULTS

Usage of AGA/AGG Codons

The positions of AGA/AGG codons in 493 different *E. coli* genes available in GenBank were analyzed. From these loci, 678 different polypeptides are produced of which DNA sequences have been determined. These 678 polypeptides can be divided into two groups on the basis of the AGA/AGG content: group I proteins (452) which have neither AGA or AGG codon, and group II proteins (226) which contain one or more than one AGA/AGG codons. The group II proteins were analyzed for the frequency of (AGA + AGG) codon usage for every 25 amino acid residues from the amino terminal ends. The frequency is defined as (n/N), where n is the total number of (AGA + AGG) codons within a unit of 25 amino acid residues, and N the total number of arginine codons within the same unit.

As shown by shaded bars in Figure 1A, AGA and AGG codons appear much more frequently within the first 25 codons. This high frequency of AGA/AGG codons in the first 25 codons becomes more dramatic when only those proteins which have only one AGA or AGG codon were analyzed from group II proteins (Figure 1B). Out of 226 group II proteins, 132 have either a single AGA or AGG codon. When these proteins (group IIa) were analyzed for the frequency of (AGA + AGG) codon usage it became more evident that either the AGA or AGG codon is preferentially used within the first 25 codons as shown by shaded bars in Figure 1B. This tendency becomes less significant for the remaining 94 proteins in group II which have more than one AGA/AGG codon (group IIb) as shown by shaded bars in Figure 1C. In Figure 1, the frequency of codon usage for every 25 codons is also expressed as a fraction of the total codons used for each group. These frequencies are indicated by slashed bars. It should be noticed that approximately 40% of the genes containing a single AGA or AGG codon (group IIa) have the minor codon within the first 25 codons (Figure 1B). This preferential usage of AGA/AGG codons in the first 25 codons is statistically significant as judged by standard deviations indicated by T bars in Figure 1.

Usage of Other Minor Codons

In addition to AGA/AGG codons, we have analyzed usages of other minor codons. Figure 2 shows frequencies of the CUA codon usage for every 25 codons. CUA is the least used codon among the codons for Leu (2.3% of total Leu codons used in *E. coli* genes; see ref. 3). The analysis was carried out exactly in the same manner as described for AGA/AGG codons in Figure 1. There are a total of 278 genes containing CUA codons (group II; Figure 2A), out of which 142 contain only one CUA (group IIa; Figure 2B), and the remaining 136 genes contain more than one CUA codons (group IIb; Figure 2C). Again one can see clear preferential usage of the Leu minor codon within the first 25 codons although it appears to be slightly less dramatic than the case of the AGA/AGG usage (Figure 1). Approximately 30%

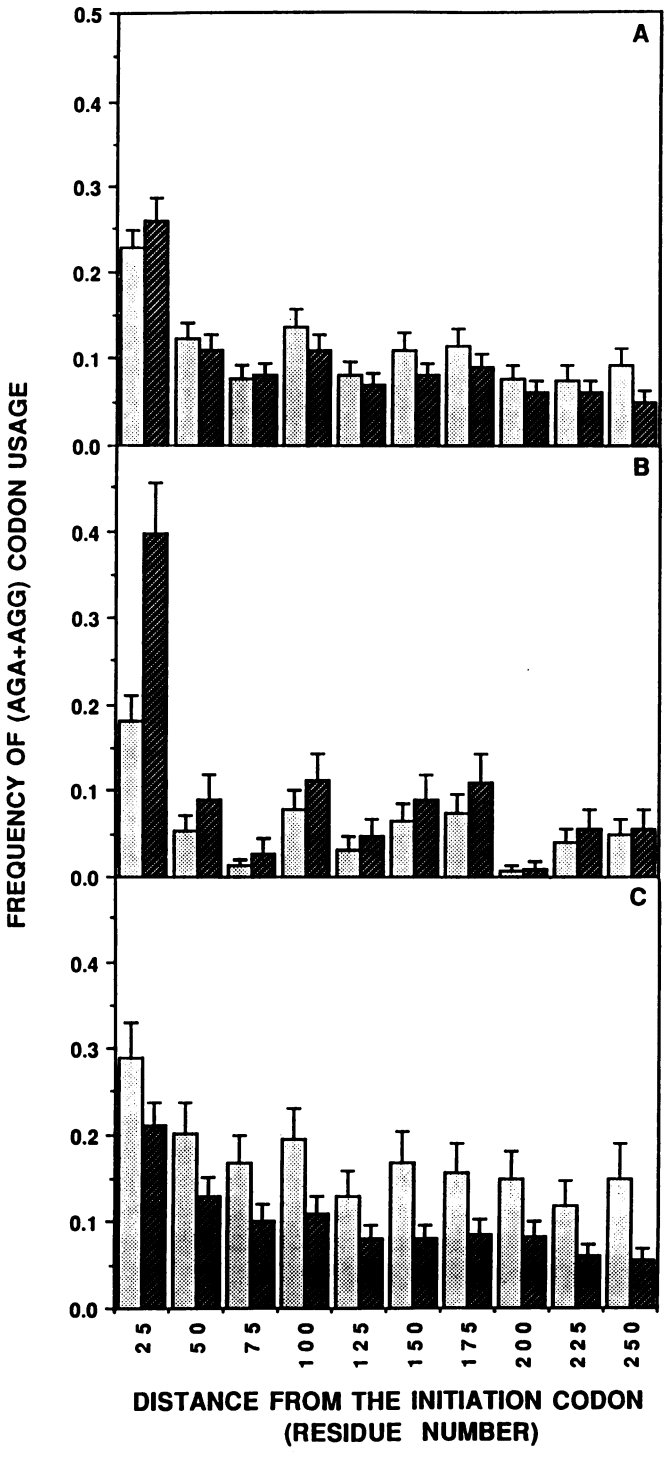


Figure 1. Frequency of AGA and AGG codon usage in *E. coli* genes. Out of 678 *E. coli* genes available in GenBank, 226 genes contain one or more AGA/AGG codons. In A, these 226 genes were analyzed for the frequency of AGA/AGG usage in sequential steps of 25 codons from the initiation codon. In B, 132 genes which have only a single minor arginine codon (AGA or AGG) were analyzed, and in C, the remaining 94 genes which have more than one AGA/AGG codon were examined. The frequencies indicated by shaded bars represent those defined as (n/N), where n is the total number of (AGA+AGG) codons within a unit of 25 codons, and N the total number of arginine codons within the same unit. The frequencies indicated by slashed bars represents those calculated as a fraction of the total AGA/AGG codons used for each group of 25 codons. T bars represent upper limit of the standard deviations.

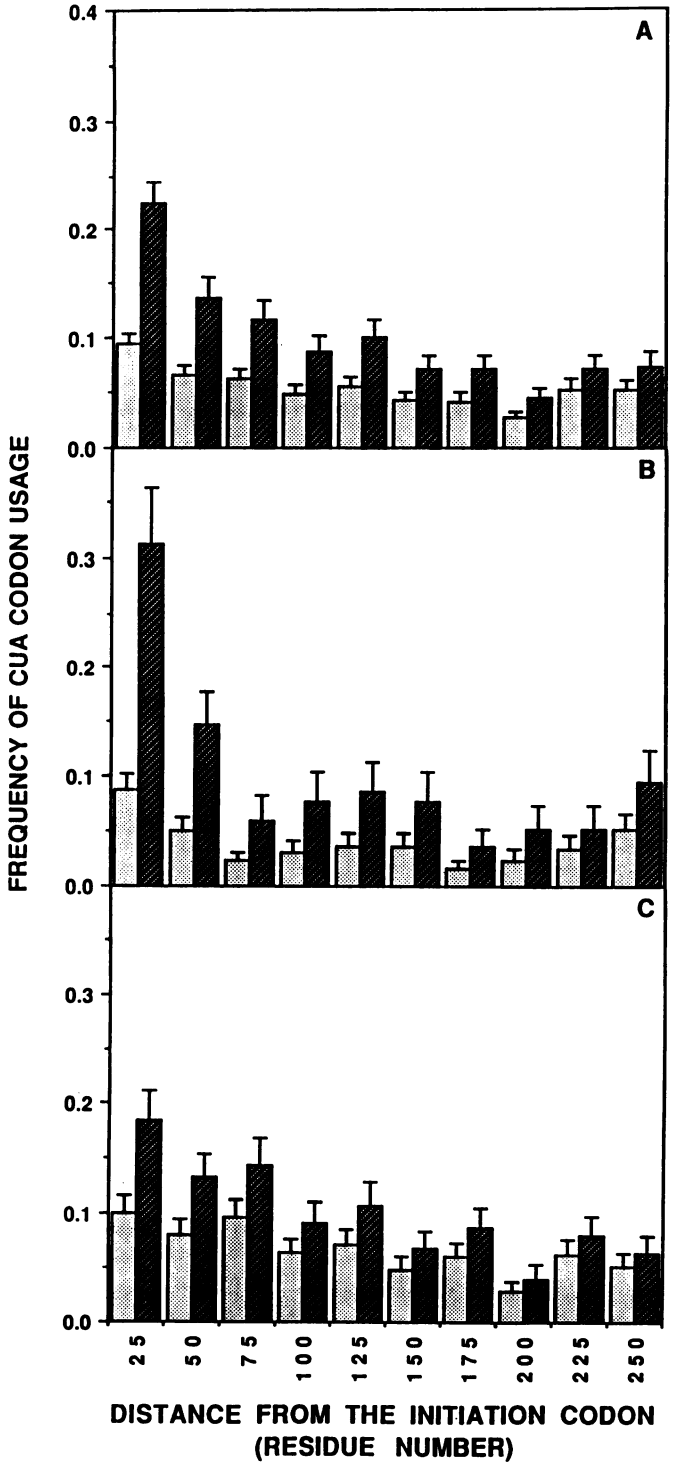


Figure 2. Frequency of CUA codon usage in *E. coli* genes. Analyses were carried out in the same way as described for AGA/AGG codon usage (Figure 1). In A, 278 genes were analyzed. In B, 142 genes containing only one CUA codon were examined and in C, the remaining 136 genes containing more than one CUA codon were analyzed.

of the genes containing a single CUA codon have the minor codon within the first 25 codons (Figure 2B).

Similar analyses were performed for other minor codons; UCA plus AGU (Ser; Figure 3-I), AUA (Ile; Figure 3-II), ACA (Thr;

Figure 3-III), CCC (Pro; Figure 3-IV) and GGA (Gly; Figure 3-V). Analyses in Figure 3 were carried out only for group II and group IIa (see above), and again clearly demonstrate preferential usages of these minor codons within the first 25 codons.

Effects of Insertion of Arginine Codons on Gene Expression

The results described above raises an intriguing possibility that the expression of some *E. coli* genes may be modulated by AGA/AGG and other minor codons. When these minor codons exist near the initiation codons, the translation initiation complex may become unstable as a result of a slow translation rate at these minor codons (10). On the basis of this hypothesis (minor codon modulator hypothesis), we next examined the effects of the arginine minor codons on gene expression by inserting one to five AGG codons after the 10th codons from the initiation codon of the *lacZ* gene. For this purpose, we used pIT-lacI containing a high expression *lacZ* system in which the *lacZ* expression is under the control of an improved *lpp* promoter (lpp^{P5}) as well as an inducible *lac* promoter-operator (16,20). In this plasmid, the first 8 amino acid residues of the original β-galactosidase were replaced by a new sequence consisting of 9 amino acid residues derived from a high expression vector, pINIII-A3 (21) as shown in Figure 4A. As a result of this replacement, three restriction sites, *Eco*RI, *Hin*dIII and *Bam*HI, are introduced in this region. It has been shown that cells harboring pIT-lacI produced β-galactosidase at a level of approximately 10% of the total cellular protein in the presence of isopropyl-β-D-galactopyranoside (IPTG) (16,20). AGG codons were inserted at the *Bam*HI site, and the sequences inserted at the *Bam*HI site (shown by an arrow in Figure 4A) are listed in Figure 4B. These sequences were added using either the unique *Hin*dIII and *Bam*HI sites or the *Bam*HI site alone, and the sequences shown in Figure 1B indicate only the extra sequences added at the site shown by the arrow in Figure 4A. As a control, five CGU codons, the major arginine codon were also inserted at the same site (see Figure 4B). The modified β-galactosidases produced from these plasmids were designated Gal-AGGN for those containing AGG codons, where N represents the number of AGG codons added. Similarly, β-galactosidase with five CGU codons was designated Gal-CGU5. It should be noted that β-galactosidase from pIT-lacI contains a total of 66 arginine residues, none of which is coded by either AGA or AGG (22).

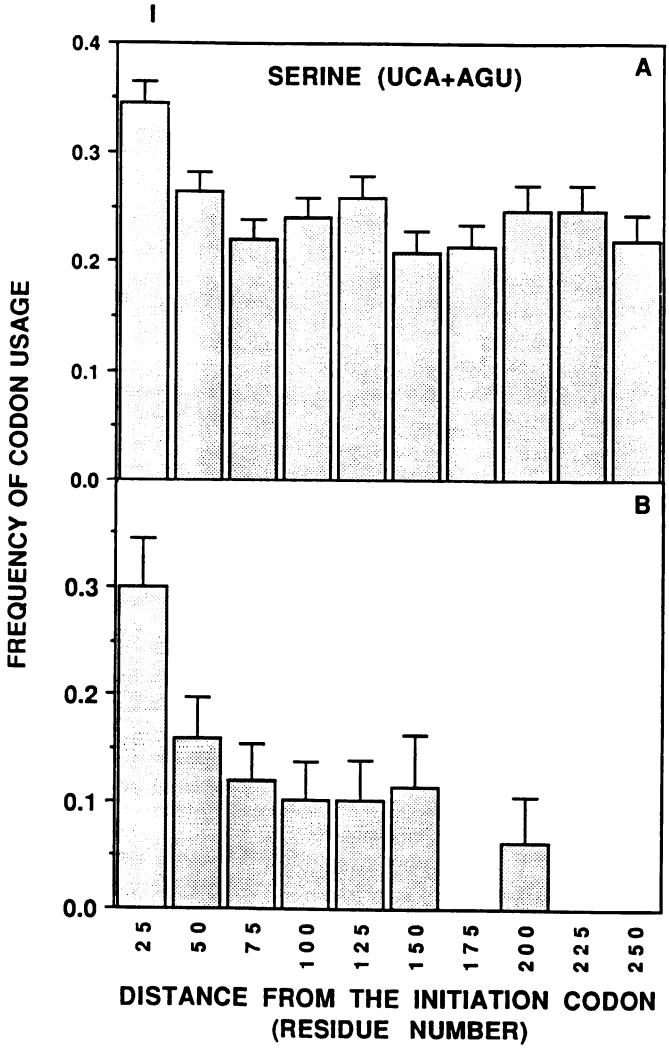
E. coli SB4288 was transformed with these modified plasmids as well as pIT-lacI and grown in M9-Casamino Acid medium. At a Klett unit of 15, IPTG was added to a final concentration of 2 mM. At time intervals after induction, aliquots were taken and total protein patterns were analyzed by SDS-polyacrylamide gel electrophoresis. The production of β-galactosidase was estimated by densitometric scanning of the gels stained by Coomasie Brilliant Blue. At 22 hr induction, the production of AGG1, AGG2, AGG3, AGG4 and AGG5 was 55, 56, 54, 32 and 11% of the production of the control β-galactosidase (no insertion; approximately 10% of total cellular protein as can be seen in lane 10, Figure 7) respectively. In contrast, the production of Gal-CGU5 was 66% of the control.

Production of Gal-AGG5 During Cell Growth

On the basis of the results described above, the Gal-AGG5 production was examined in more detail during cell growth and compared with the productions of Gal-CGU5 and the control β-galactosidase.

Figure 5 shows growth curves and β-galactosidase production after the addition of IPTG to cells carrying pIT-lacI (control), pIT-lacI-CGU5 and pIT-lacI-AGG5. There are no discernible differences in the growth profiles of these cells (Figure 5A). However, when β-galactosidase production was estimated by densitometric assay of the stained gels after SDS-polyacrylamide gel electrophoresis of total cellular protein, one can observe substantial differences among them (Figure 5B). During the first 3 hrs after induction, both Gal-CGU5 and Gal-AGG5 were produced at the same rate (approximately 50% of the rate of the control β-galactosidase production). After 3 hr induction, however, the Gal-AGG5 production completely ceased, while the Gal-CGU5 production continued. For both the control and Gal-CGU5, production did not stop even in stationary phase (Figure 5B). As a result, after 22 hr induction, the amount of β-galactosidase in the control cells reached approximately 10% of total cellular protein, and the amount of Gal-CGU5 was 66% of the control, while the amount of Gal-AGG5 was only 11% of the control.

These results indicate that there are two distinct phases for the Gal-AGG5 production during cell growth, phase I and phase II. Phase I corresponds to early growth phases up to mid-log phase, and during this phase Gal-AGG5 is produced as efficiently as Gal-CGU5. However, after mid-log phase (phase II) the Gal-



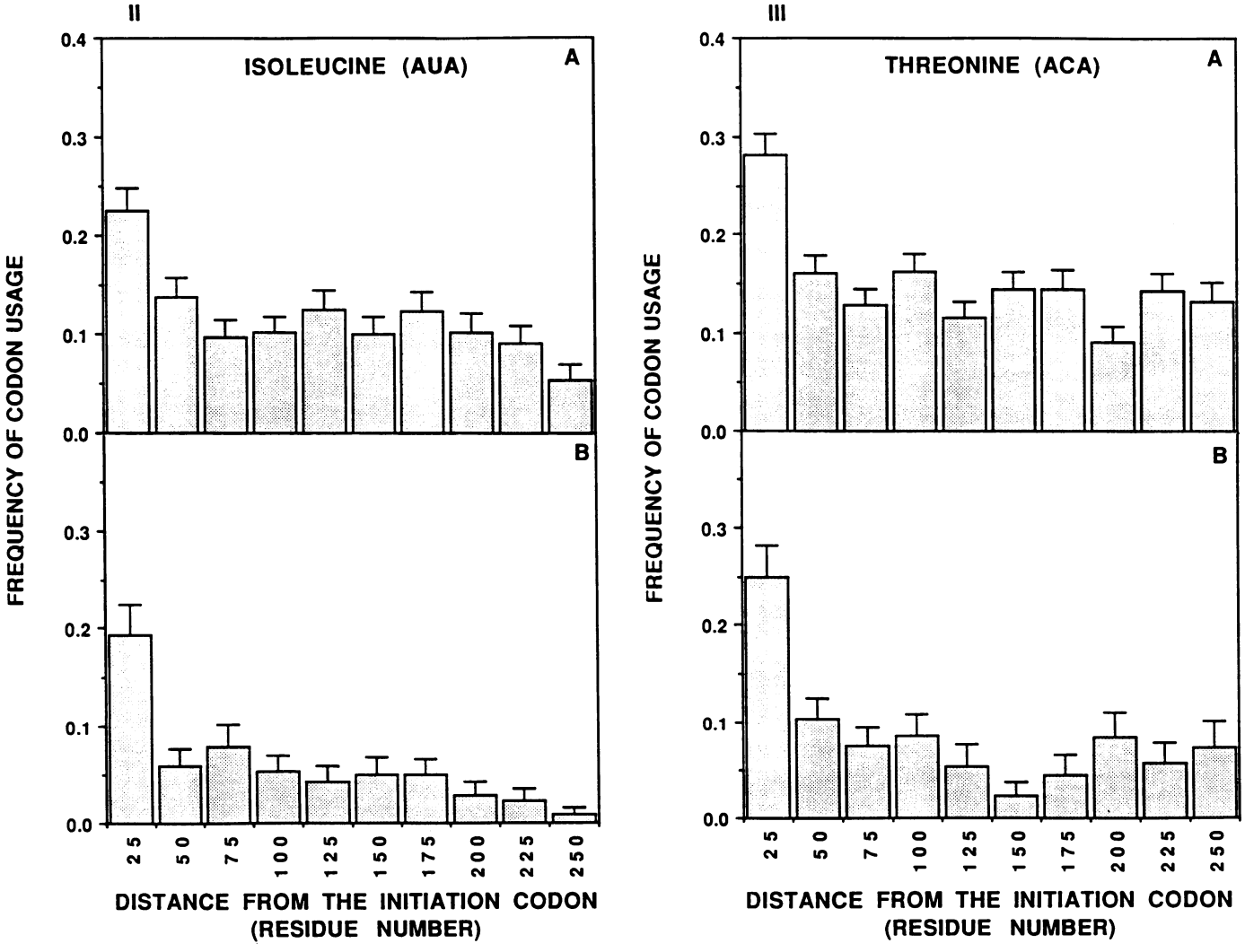
AGG5 production completely stops, while the Gal-CGU5 production continues. The results described above were obtained using a *relA* strain. However, the identical results were obtained using *E. coli* JM83 (*relA*$^+$) (data not shown), indicating that the present results are not associated with the *relA* mutation. It is interesting to note that in contrast to Gal-AGG5, the production of Gal-AGG4 continued during the first 9 hrs of induction and then stopped completely (data not shown). Furthermore, there were no significant differences in the β-galactosidase production among Gal-AGG1, Gal-AGG2 and Gal-AGG3, which continued for 22 hrs after induction at the approximately 50% level of the control (pIT-lac1) (data not shown).

Suppression of the Negative Effect of Five AGG Codon Insertion

The dramatic negative effect of insertion of five AGG codons in the *lacZ* gene on the β-galactosidase production in phase II is most likely due to the limited availability of charged tRNA for AGG. However, it is not certain whether the extremely low concentration of charged tRNA for AGG has a negative effect on all the genes containing the minor arginine codons. In view of the fact that the AGA/AGG codons are preferentially found in the first 25 codons as discussed earlier, it is quite possible that the position of the AGG insertion site in a gene may have some effect on gene expression.

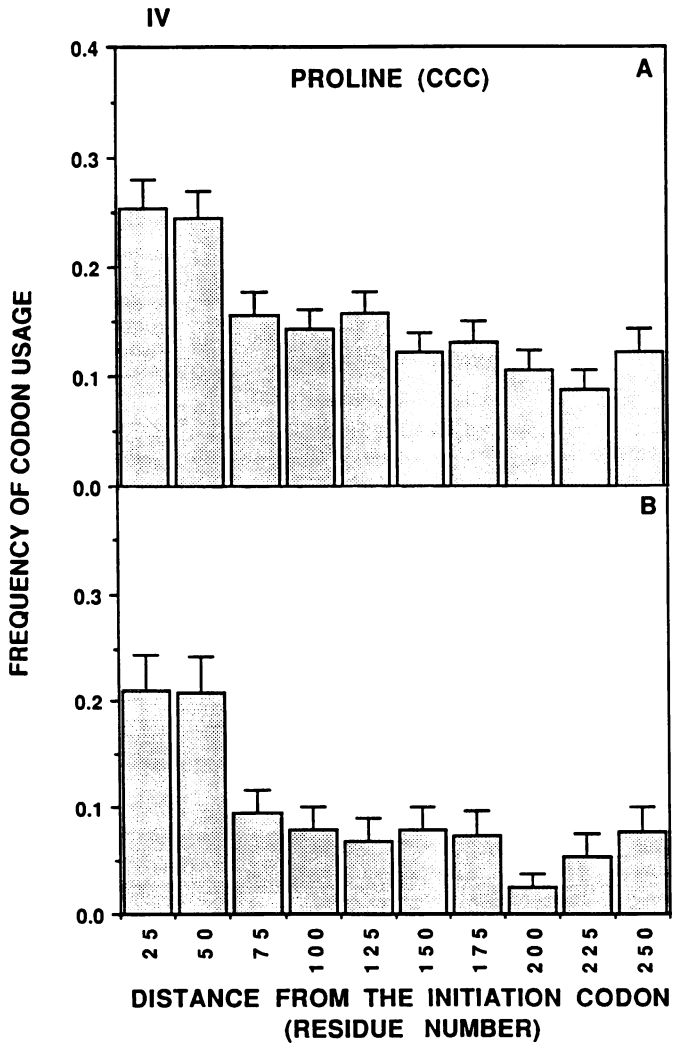
In fact, Robinson et al. (23) inserted three AGG codons into the middle of the gene for chloramphenicol acetyltransferase (CAT), and did not observe any significant effect of the minor AGG codons on production of CAT. It was even found to be produced at a rate of as high as 16% of the total cellular protein. One major difference between this experiment and ours was the site of insertion of the minor codons. For CAT the AGG was inserted after the 72nd codon from the initiation codon whereas in the Gal-AGG5 construction in the present study, the AGG codons were placed after the 10th codon following the initiation codon. Thus it is reasonable to speculate that the distance between the initiation codon and the AGG insertion site may play a crucial role in affecting gene expression.

In order to test this hypothesis, 18-mer oligonucleotides (see Figure 4C) were inserted at the unique *Hin*dIII site 5 amino acid residues after the initiation methionine residue as shown in Figures 6. Annealing of two 18-mer oligonucleotides generated a short fragment with *Hin*dIII compatible 4-base 5′ overhangs at both ends. Insertion of a monomer of the annealed oligonucleotides adds 6 amino acid residues between residue-6 and residue-7 (see Figure 4C). In orientation 1, Ser-Ser-Gly-Ala-Gln-Ala is inserted, each residue being coded for by major



codons. In the opposite orientation, orientation 2, Ser-Leu-Pro-Val-His-Gln is added, each residue again coded for by major codons, except in the cases of TTG for Leu and CCT for Pro. However, TTG and CCT codons are much more frequently used than AGA and AGG codons. From the analysis of 199 *E. coli* genes listed in the 1986 GenBank database (3), 10.1% of total Leu codons used are TTG and 12.5% of total Pro codons are CCT in contrast to AGA (1.7%) and AGG (0.9%). In order to facilitate multimer insertion at the *Hin*dIII site, the annealed oligonucleotides were ligated by themselves before ligating into the plasmid *Hin*dIII site. Resulting transformants were selected by color development of McConkey plates. Those colonies which developed denser red color than the control colony with pIT-lacI-AGG5 were isolated and sizes of inserted fragments were analyzed after digesting the isolated plasmid DNA with *Xba*I and *Bam*HI. The exact sizes and sequences of the inserted fragments were determined by DNA sequencing. By this method, 2 monomers (orientation 1 and 2), 2 dimers (1-1 and 2-1), 1 trimer (2-2-1), 2 tetramers (1-1-2-1 and 2-1-2-2), 2 hexamers (2-1-2-2-2-1 and 2-1-2-2-2-2), 1 heptamer (1-2-1-1-1-1-2), 1 octamer (2-1-1-1-2-1-1-2) and 1 nonadecamer (2-1-1-1-1-2-1-1-1-1-2) were isolated.

The β-galactosidase production was analyzed after overnight culture of the cells harboring the plasmids isolated above by SDS-polyacrylamide gel electrophoresis. An example of the analyses is shown in Figure 7, which clearly demonstrates that as the sizes of the inserted spacer fragments increased, β-galactosidase production was dramatically enhanced. In Figure 7, monomer (1), dimer (1-1), trimer (2-2-1), tetramer (1-1-2-1), hexamer (2-1-1-2-2-1), heptamer (1-2-1-1-1-1-2), octamer (2-1-1-1-2-1-1-2) were analyzed. Other polymer insertion constructions were also analyzed by SDS-polyacrylamide gel electrophoresis and β-galactosidase production was estimated by the density of the band. The relative production of all the polymer insertion constructions were plotted against the distances from the initiation codon. As shown in Figure 8, the β-galactosidase production increased almost linearly up to 8 fold as the distance between the initiation codon and the site of AGG codons increased. Although the orientation of the inserted spacer fragments showed some effects on the production of β-galactosidase, the distance from the initiation codon appears to



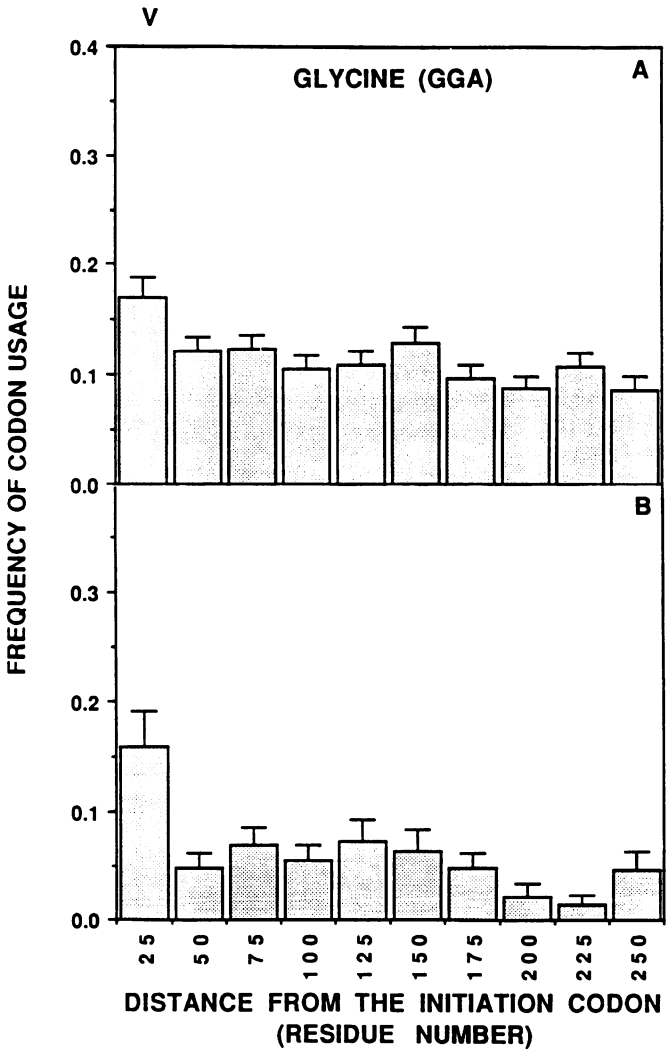


Figure 3. Frequency of other minor codon usage in *E. coli* genes. The analyses were performed in the same manner as described for AGA/AGG codon usage (Figure 1) except that only frequencies calculated as n/N (see the legend for Figure 1) are shown. I, UCA plus AGU for Ser. 520 genes for A and 107 genes for B; II, AUA for Ile. 241 genes for A and 124 genes for B; III, ACA for Thr. 419 genes for A and 158 genes for B; IV, CCC for Pro. 350 genes for A and 177 genes for B, and V, GGA for Gly. 401 genes for A and 138 genes for B.

be the major factor determining the yield of the enzyme. When the distance became 76 amino acid residues, the yield approached the level obtained with the control gene lacking any arginine codon insertion. The identical dramatic enhancement of yield by the spacer insertion was also observed with a *relA*$^+$ *E. coli* strain (strain JM83; data not shown), again indicating that these phenomena are not associated with the *relA* mutation of *E. coli* SB4288.

DISCUSSION

AGA/AGG Modulator Hypothesis (Minor Codon Modulator Hypothesis)

We found that in exponentially growing cells minor arginine codons have no negative effects on gene expression. However, once cell growth passes mid-log phase, the minor arginine codons start to exhibit severe negative effects on the expression of a gene containing AGA/AGG codons, particularly when these minor codons are located in a close proximity to the initiation codon.

We propose the following hypothesis (AGA/AGG modulator hypothesis or minor codon modulator hypothesis) to explain the regulatory mechanism by AGA/AGG codons. First it is assumed that the gene for tRNA for AGG and AGG is under very tight transcriptional control so that after mid-log phase of cell growth the availability of charged tRNA for AGA and AGG becomes very limited. This assumption is consistent with the fact that tRNA for AGA and AGG is coded by the *dnaY* gene, which is associated with the regulation of DNA replication (13). As a result of the

Figure 4. DNA Sequences and Amino Acid Sequences of a Part of β-Galactosidase and Oligonucleotides Used. A, the DNA sequence for the first 8 amino-terminal amino acid residues of β-galactosidase (16). This sequence was changed in pIT-lacI-β-galactosidase to the sequence shown below, which contained *Eco*RI (E), *Hin*dIII(H), and *Bam*HI(B) sites. B, oligonucleotides inserted at the *Bam*HI site of pIT-lacI shown by a big arrow. C, the linker inserted at the *Hin*dIII site of pIT-lacI-AGG5. Two 18-mer oligonucleotides, $^{5'}$AGCTCTGGTGCACAGGCA$^{3'}$ and $^{5'}$AGCTTGCCTGTGCACCAG$^{3'}$ are able to anneal to form two identical 5'-end overhangs. Thus, the annealed oligonucleotides can be inserted in two different orientations as shown in the *Hin*dIII site.

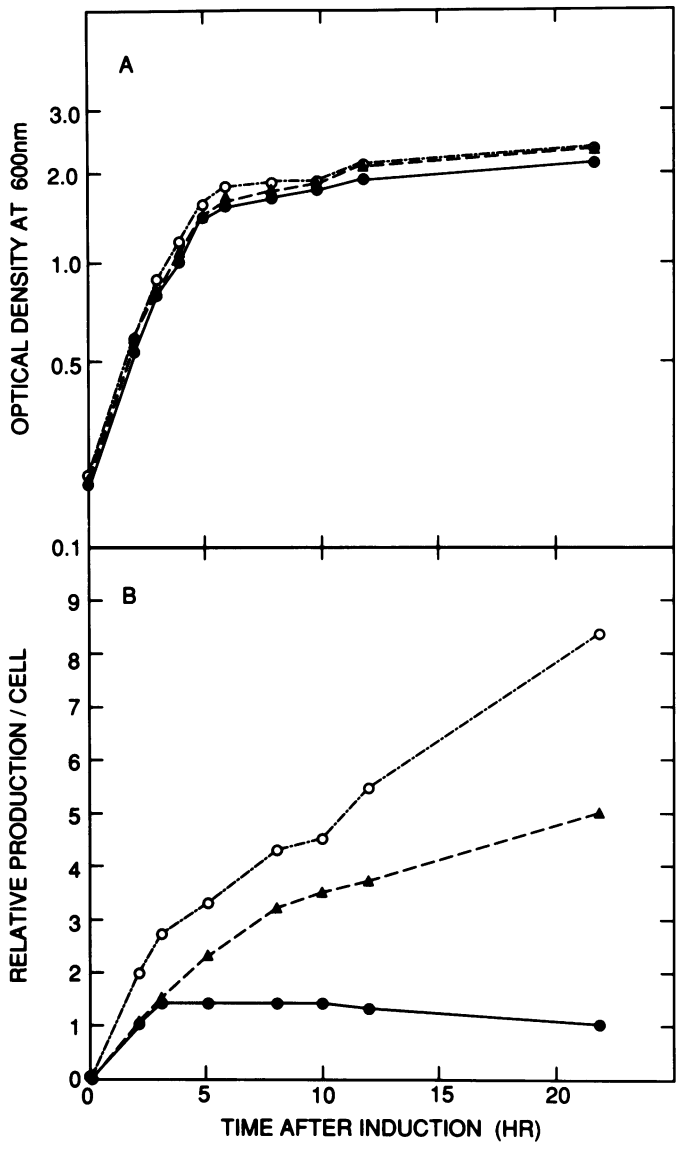


Figure 5. Growth curves and β-Galactosidase Production of Cells Harboring pIT-lacI-AGG5, pIT-lacI-CGU5 and pIT-lacI. *E. coli* SB4288 harboring the minor Arg codon plasmid pIT-lacIA-AGG5, the major Arg codon plasmid pIT-lacI-CGU5 or a control parent plasmid pIT-lacI were grown as described in 'Materials and Methods'. After induction of *lacZ* expression with 2mM IPTG, aliquots were taken at the times indicated. Cells were collected by centrifugation and washed once with 20mM sodium phosphate buffer (pH 7.1). To the pellet, 0.1 ml of solubilizing solution (2% SDS and 5% glycerol in 125 mM Tris-HCl buffer, pH 6.8) was added, and proteins were solubilized by incubating the mixtures in a boiling water bath for 3 min. To each lane, an amount of sample equivalent to 0.2 OD_{600} units (approximately 1.5×10^7 cells) was applied. SDS-polyacrylamide gel electrophoresis was carried out using a 10% polyacrylamide (25). The gel was subsequently stained with Coomassie Brilliant Blue.

A, growth curves after the addition of 2mM IPTG at 0 time. Growth was measured by a Klett-Summerson colorimeter. B, β-galactosidase production after the addition of 2mM IPTG at 0 time. The amounts of β-galactosidase were estimated by densitometric scanning of the Coomassie Brilliant Blue stained β-galactosidase bands. The amounts were expressed as ratios, relative to the density of β-galactosidase of pIT-lacI at 22 hr after induction. ●——●, pIT-lacI-AGG5; ▲- - -▲, pIT-lacI-CGU5, and ; ○- - -○, pIT-lacI.

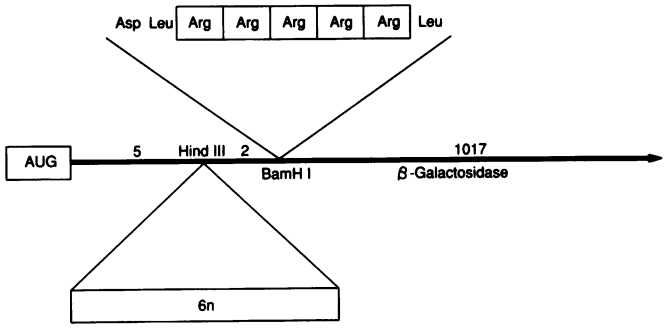


Figure 6. Insertion of Oligonucleotides into the β-Galactosidase Gene of pIT-lac1. The insertion of oligonucleotide AGG5 (Figure 4B) at the *Bam*HI site adds a sequence of 8 amino acid residues containing 5 contiguous arginine residues after the 8th residue from the initiation codon AUG. The insertion of a multimer of the linker shown in Figure 4C at the *Hin*dIII site adds a sequence of 6×n amino acid residues (n=the number of the linker monomers) 5 amino acid residues after the initiation codon, AUG. The numbers in the figure indicate the numbers of amino acid residues.

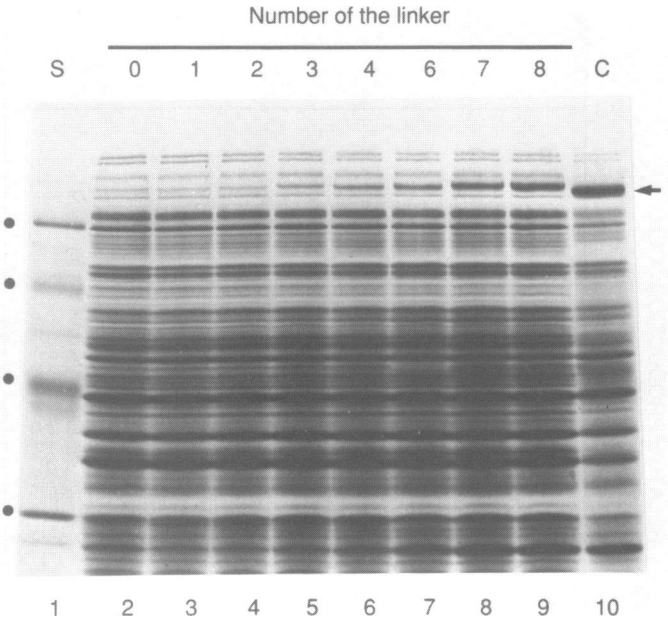


Figure 7. SDS-Polyacrylamide Gel Electrophoresis of β-Galactosidase with Various Spacing Between the Initiator AUG and AGG Codons. Cells harboring pIT-lac1A with various numbers of the spacer linker fragments were grown for 22 hr after the addition of 2mM IPTG. SDS-polyacrylamide gel electrophoresis was carried out as described in 'Materials and Methods'. The numbers of the linker inserted at the *Hin*dIII site of pIT-lac1-AGG5 (see Figures 4 and 6) are indicated on the top of the gel. Lane 1, molecular weight standards. The dots from the top indicate the positions of marker proteins, 92.5, 66, 45, and 31 kD, respectively; lane 2, cells harboring pIT-lac1-AGG5; lane 3, pIT-lac1-AGG5 with one linker (orientation 1, see Figure 4); lane 4, with two linkers (1-1); lane 5, with three linkers (2-2-1); lane 6, with four linkers (1-1-2-1); lane 7, six linkers (2-1-1-2-2-1); lane 8, with seven linkers (1-2-1-1-1-1-2); lane 9, with eight linkers (2-1-1-1-2-1-1-2), and lane 10, pIT-lac1 as a control. An arrow indicates the position of β-galactosidase produced by pIT-lac1 (lane 10). Note the samples of β-galactosidases in other lanes migrated slower than that in lane 10 due to increased size from the linker insertions.

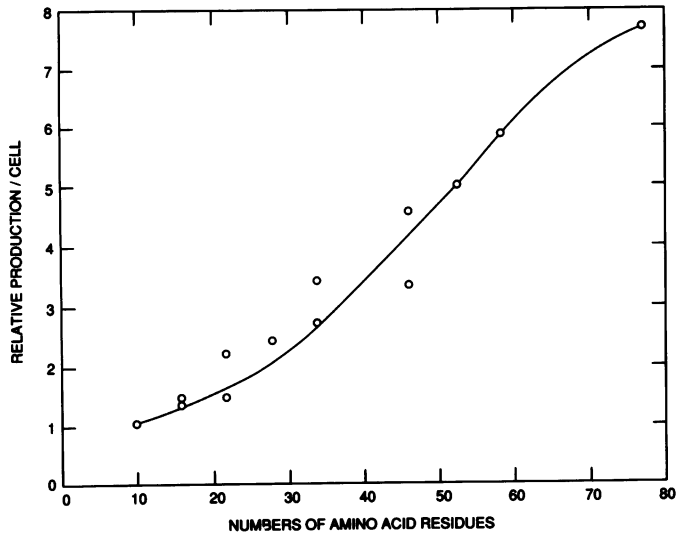


Figure 8. Relative Production of β-Galactosidases by pIT-lac1-AGG5 with Various Numbers of Spacer Fragments. The production of β-galactosidase by pIT-lac1-AGG5 with various numbers of the spacer linker fragments were estimated after 22 hr of induction. The production was expressed as relative ratios of pIT-lac1-AGG5 with linker to that of pIT-lac-AGG5 without the linker insertion, and plotted against the number of the amino acid residues from the initiation methionine to the first arginine residue coded by AGG. Thus, the point at 10 residues corresponds to pIT-lac1-AGG5 with no insertion, two points at 16 residues to monomer insertion (orientation 2 and 1 for the upper and the lower points, respectively), two points at 22 residues to dimer insertion (1-1 and 2-1), one point at 28 residues to trimer insertion (2-2-1), two points at 34 residues to tetramer insertion (2-1-2-2 and 1-1-2-1), two points at 46 residues to hexamer insertion (2-1-2-2-2-2 and 2-1-2-2-2-1), one point at 52 residues to heptamer insertion (1-2-1-1-1-1-2), one point at 58 residues to octamer insertion (2-1-1-1-2-1-1-2) and one point at 76 residues to nonadecamer (2-1-1-1-1-2-1-1-1-1-2). The production of β-galactosidase with monomer (1), dimer (1-1), trimer (2-2-1), tetramer (1-1-2-1), hexamer (2-1-1-2-2-1), heptamer (1-2-1-1-1-1-2) and octamer (2-1-1-1-2-1-1-2) insertions were from Figure 7. All the other points were estimated from another gel run in the same manner as Figure 6 (not shown).

limited availability of tRNA for AGA and AGG, translation of certain mRNAs containing AGA/AGG codons near the initiation codon is severely hampered. A ribosome translating such an mRNA has to pause at the minor arginine codons until a charged tRNA for AGA and AGG becomes available. On the basis of the recent report that the translation rate slows down at minor codons (10), a ribosome is considered to move slowly at the region containing the AGG codons. Our results thus can best be explained if one assumes that the paused ribosome—mRNA complex becomes unstable or the rate of the formation of the initiation complex decreases when the ribosome pauses close to the initiation codon, and that its stability or the rate of initiation increases linearly as the distance between the initiation codon and the minor codon increases. When this distance becomes greater than 50 to 60 codons the paused complex becomes stable so that no more ribosomes fall off at the minor arginine codons. It is important to note that AGA/AGG codons are preferentially used in the first 25 codons in *E. coli* genes particularly if a gene contains only one AGA or AGG (Figure 1). This fact together with the present results strongly support the important role of the minor codons in gene expression when they exist within the first 20 to 25 codons of a gene.

The results shown in Figure 8 indicates that the stability of the paused ribosome—mRNA complex may be directly correlated to the length of a nascent peptide; the shorter the nascent peptide, the less stable the complex, so translation is more likely to terminate at the AGA/AGG codon. Thus, the AGA/AGG codon functions as a modulator for translation when the availability of charged tRNA for AGA/AGG becomes limited. It should be pointed out that the primary effect of the AGG codons described

here is unlikely to be due to the stability of mRNA, since all the mRNAs containing different distances between the initiation codon and the AGG codons have exactly the same mRNA structure after the AGG codons. Therefore, if the pause of a ribosome at the AGG codon site causes destabilization of the mRNA downstream of the AGG codon site, the effect is assumed to be identical in all cases. This further supports our hypothesis that the minor codons modulate gene expression by regulating the rate of initiation complex formation depending on the distance between the initiation codon and the minor codons, but not by regulating the rate of peptide elongation. This notion is supported by the recent results on the estimation of the translation rates of minor and major codons (10). It has been reported that in the coat protein gene of phage MS2 ribosomal frameshifting occurs at the sequence AGG-AGG (24). However, the present suppression effect cannot be explained by frameshifting, because such ribosome frameshifts prematurely terminate polypeptide synthesis. Note that the increase of the sizes of β-galactosidases in Figure 7 correlate well with the sizes of the spaces inserted. The effect of the relative position of minor codons in an mRNA on protein synthesis has been discussed, and it was predicted that tandem repeats of minor codons dramatically reduce the maximum level of protein synthesis (26). However, our results indicate that the most important factor for protein synthesis is the distance between the initiation codon and the site of minor codons, but not clustering of minor codons.

General Effects of Minor Codons on Gene Expression

Among the *E. coli* minor codons, the AGA and AGG codons appear to be unique in two aspects: (1) they are the least used codons among the minor codons, and (2) the tRNA for AGA and AGG is the gene product of *dnaY*. The proposed AGA/AGG modulator hypothesis presents a novel mechanism for global regulation of cellular functions. During nutrient deprivation the AGA/AGG codons function as a modulator to inhibit the production of key proteins in *E. coli* so that various cellular activities such as DNA synthesis cease in a well balanced fashion. Under such nutrient limited conditions, the distance between the initiation and the AGA/AGG codons subtly affects the synthesis of key functions which regulate global cellular activities. It is interesting to note that among a group of the proteins which have an AGA or AGG codon within the first 25 codons, there are various essential gene products such as (a) a protein required for DNA replication [single-strand DNA-binding protein (4th position)], (b) proteins associated with gene regulation [adenylate cyclase (11th position), LexA protein (7th position), and IHF (9th position)], (c) proteins associated with protein synthesis [glycyl-tRNA synthetase α subunit (7th position), phenylalanyl-tRNA synthetase β subunit (2nd position), S10 (5th position)], and (d) other essential proteins [FtsA protein (7th position), and thioredoxin precursor (11th position)]. However, it should be noted that in the presence of pIT-lacI system, significant reduction of β-galactosidase production in stationary phase was observed only when more than three AGG codons were inserted in the *lacZ* gene. It is possible that there are other factors in addition to the arginine minor codons such as other codons near by the arginine codons which synergistically affect the efficiency of mRNA translation.

It should be noted that in addition to AGA/AGG codons such as CUA (Leu), UCA (Ser), AGU (Ser), AUA (Ile), ACA (Thr), GGA (Gly) and Pro (CCC) are also used preferentially within the first 25 codons in particular when they are used only once within a gene. Therefore, it is tempting to think that not only AGA/AGG codons but also other minor codons are able to modulate gene expression by placing them in an earlier part of a gene. This regulatory mechanism of gene expression by minor codons is considered to become functional only when cells are starved of amino acids, thus playing an important role in global regulation of cellular activities.

ACKNOWLEDGEMENT

The authors are grateful to Drs. Danny Reinberg, Claude Lerner and Steve Pollitt for their critical reading of the manuscript. We would also like to thank Dr. C.T. Chen for computer programming to analyze the *E. coli* gene products.

The work was supported by grants from U.S. Public Health Service (GM 19043) and Takara Shuzo Co. Ltd.

REFERENCES

1. Ikemura, T. (1981a) *J. Mol. Biol.* **146**, 1–21.
2. Ikemura, T. (1981b) *J. Mol. Biol.* **151**, 389–409.
3. Maruyama, T., Gojobori, T., Aota, S.I. and Ikemura, T. (1986) *Nucleic Acids Res.* **14**, r151-r197.
4. Nakamura, K. and Inouye, M. (1979) *Cell* **18**, 1109–1117.
5. Nakamura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye, M. (1980) *J. Biol. Chem.* **255**, 210–216.
6. Post, L.E., Stycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) *Proc. Nat. Acad. Sci.*, USA **76**, 1697–1701.
7. Nomura, M., Yates, J.L., Dean, D. and Post, L.E. (1980) Proc. Nat. Acad. Sci., USA, **77**, 7084–7088.
8. Holm, L. (1986) *Nucl. Acids Res.* **14**, 3075–3087.
9. Sharp, P.M. and Li, W-H. (1986) *Nucl. Acids Res.* **14**, 7737–7749.
10. Sorensen, M.A., Kurland, C.G. and Pedersen, S. (1989) *J. Mol. Biol.* **207**, 365–377.
11. Kurland, C.G. (1987) *Trends Biochem.* **12**, 126–128.
12. Hoekema, A., Kastelein, R.A., Vasser, M. and de Boer, H.A. (1987) *Mol. Cell Biol.* **7**, 2914–2924.
13. Garcia, G.M., Mar, P.K., Mullin, D.A., Walker, J.R. and Prather, N.E. (1986) *Cell* **45**, 453–459.
14. Green, P.J. and Inouye, M. (1984) *J. Mol. Biol.* **176**, 431–442.
15. Vieira, J. and Messing, J. (1982) *Gene* **19**, 259–268.
16. Inouye, S. and Inouye, M. (1985) *Nucl. Acids Res.* **13**, 3101–3110.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* USA **74**, 5463–5467.
18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: a laboratory manual*, Cold Spring Harbor, New York.
19. Miller, J.H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor, New York.
20. Tsung, K., Inouye, S. and Inouye, M. (1989) *J. Biol. Chem.* **264**, 10104–10109.
21. Masui, Y., Coleman, J. and Inouye, M. (1983) *Experimental Manipulation of Gene Expressions* (ed. M. Inouye) pp 15–32, Academic Press, New York.
22. Kalnins, A., Otto, K., Ruther, U. and Muller-Hill, B. (1983) *EMBO J.* **2**, 593–597
23. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) *Nucl. Acids Res.* **12**, 6663–6671.
24. Spanjaard, R.A. and van Duin, J. (1988) *Proc. Natl. Acad. Sci.* USA **85**, 7967–7971.
25. Anderson, C.W., Baum, P.R. and Gestland, R.F. (1973) *J. Virol.* **12**, 241–252.
26. Varenne, S. and Lazdunski, C. (1986) *J. Theor. Biol.* **120**, 99–110.